Matrix Decomposition

This chapter is about three aspects of matrix:

- how to summarise matrices
- how matrices can be decomposed
- how these decompositions can be used for matrix approximations

Determinant and Trace

Determinant

A determinant is a mathematical object in the analysis and solution of systems of linear equations. Determinants are *only* defined for square matrices $A \in \mathbb{R}^{m \times n}$.

> 💡 In this book, we write determinant as $\det(A)$ or sometimes as $|A|$.

The determinant of a square matrix $A \in \mathbb{R}^{m \times n}$ is a function that maps A onto a real number.

Theorem 4.1. For any square matrix $A \in \mathbb{R}^{n \times n}$ it holds that A is invertible if and only if $\det(A) \neq 0$.

Closed-form expressions for determinants of small matrices in terms of the elements of the matrix are:

- For $n = 1$,

$$\det(A) = \det(a_{11}) = a_{11}$$

- For $n = 2$,

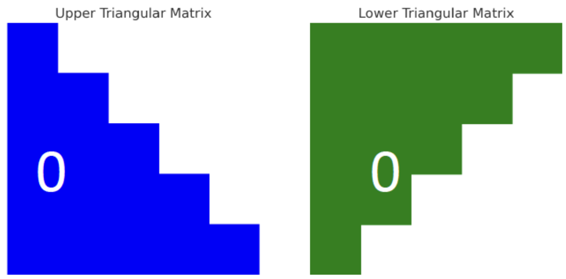
$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

- For $n = 3$ (known as `Sarrus' rule`),

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{33}a_{12}a_{21}$$

For a square matrix T

- if $T_{ij} = 0$ for $i > j$, i.e., the matrix is zero below its diagonal, then T is called *upper-triangular matrix*
- if $T_{ij} = 0$ for $i < j$, i.e., the matrix is zero *above* its diagonal, then T is called *lower-triangular matrix*



For a triangular matrix (both upper and lower), the determinant is the product of the diagonal elements, i.e.,

$$\det(T) = \prod_{i=1}^{n} T_{ii}$$

To compute the determinant of an $n \times n$ matrix, theorem 4.2 introduces a more general algorithm which is recursively applied to reduce the computation of $n \times n$ matrix to $(n-1) \times (n-1)$matrix and eventually to the 2×2 matrix.

Theorem 4.2. (Laplace Expansion). Consider a matrix $A \in \mathbb{R}^{n \times n}$. Then, for all $j = 1, \ldots, n$:

1. Expansion along column j

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(A_{k,j})$$

2. Expansion along row j

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(A_{j,k})$$

Here $A_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the submatrix of A that deleting row k and column j.

> **Example**
>
> To compute the determinant of
>
> $$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$
>
> using the Laplace expansion along the first row:
>
> $$\begin{vmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix} = (-1)^{1+1} \cdot 1 \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+2} \cdot 2 \begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3} \cdot \begin{vmatrix} 3 & 1 \\ 0 & 0 \end{vmatrix}$$

> The compute the determinants of all 2×2 matrices and obtain
>
> $$\det(A) = 1(1-0) - 2(3-0) + 3(0-0) = -5$$

For $A \in \mathbb{R}^{n \times n}$ the determinant exhibits the following **properties**:

- $\det(AB) = \det(A)\det(B)$
- $\det(A) = \det(A^T)$
- If A is regular (invertible), then $\det(A^{-1}) = \frac{1}{\det(A)}$
- Similar matrices possess the same determinant. For a linear mapping $\Phi : V \to V$ all transformation matrices A_Φ of Φ have the same determinant. Thus the determinant is invariant to the choice of basis of a linear mapping.
- Adding a multiple of a column/row to another one does not change $\det(A)$
- $\det(\lambda A) = \lambda^n \det(A)$
- Swapping 2 rows/columns changes the sign of $\det(A)$

Using Gaussian elimination to bring A to a row-echelon form helps compute the $\det(A)$.

Theorem 4.3. A square matrix $A \in \mathbb{R}^{n \times n}$ has $\det(A) \neq 0$ if and only if $\text{rk}(A) = n$. In other words, A is invertible if and only if it's full rank.

> 💡 However, contemporary approaches in ML use direct numerical methods to replace the calculation of the determinant, e.g., we use Gaussian elimination to compute the inverse matrices.

Trace

Definition 4.4 The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is defined as

$$\text{tr}(A) := \sum_{i=1}^{n} a_{ii}$$

i.e., the trace is the sum of the diagonal elements of A.

Trace has the following properties:

- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$ for $A, B \in \mathbb{R}^{n \times n}$
- $\text{tr}(\alpha A) = \alpha \text{tr}(A),\ \alpha \in \mathbb{R}$
- $\text{tr}(\mathbf{I}_n) = n$
- $\text{tr}(AB) = \text{tr}(BA)$ for $A \in \mathbb{R}^{n \times k},\ B \in \mathbb{R}^{k \times n}$

The properties of the trace of the matrix products are more general. Specifically, the trace is invariant under cyclic permutations, i.e,

$$\text{tr}(AKL) = \text{tr}(KLA)$$

for $A \in \mathbb{R}^{a \times k},\ K \in \mathbb{R}^{k \times l},\ L \in \mathbb{R}^{l \times a}$.

Eigenvalues and Eigenvectors

Every linear mapping has a unique transformation matrix given an ordered basis. The eigenvalues of a linear mapping tell how a special set of vectors, the eigenvectors, is transformed by the linear mapping.

Eigenvalues

Definition 4.6. Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. Then $\lambda \in \mathbb{R}$ is an *eigenvalue* of A and $x \in \mathbb{R}^n \setminus \{0\}$ is the corresponding *eigenvector* of A if

$$Ax = \lambda x$$

This equation is called eigenvalue equation.

Some equivalent statements:

- λ is an eigenvalue of $A \in \mathbb{R}^{n\times n}$
- $\mathrm{rk}(A - \lambda I_n) < n$
- $\det(A - \lambda I_n) = 0$
- There exists an $x \in \mathbb{R}^n \setminus \{0\}$ with $Ax = \lambda x$, or, $(A - \lambda I_n)x = 0$ can be solved non-trivially, i.e, $x \neq 0$

Definition 4.7 (Collinearity and Codirection). If x is an eigenvector of A associated with eigenvalue λ, then for any $c \in \mathbb{R} \setminus \{0\}$ it holds that cx is an eigenvector of A with the same eigenvalue since

$$A(cx) = cAx = c\lambda x = \lambda(cx)$$

Thus, all vectors that are collinear to x are also eigenvectors of A.

Theorem 4.8. $\lambda \in \mathbb{R}$ is an *eigenvalue* of $A \in \mathbb{R}^{n\times n}$ if and only if λ is *a root of characteristic polynomial* $p_A(\lambda)$ of A.

> For $\lambda \in \mathbb{R}$ and a square matrix $A \in \mathbb{R}^{n\times n}$
>
> $$p_A(\lambda) := \det(A - \lambda I_n) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n$$
>
> $c_0, \cdots, c_{n-1} \in \mathbb{R}$, is the *characteristic polynomial* of A.
>
> For example, for $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, it's characteristic polynomial is
>
> $$p_A(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} = (1-\lambda)(3-\lambda) - 8 = (\lambda - 5)(\lambda + 1)$$
>
> Thus, eigenvalues are $\lambda_1 = 5, \lambda_2 = -1$.

Theorem 4.9. Let a square matrix A have an eigenvalue λ_i. The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.

Theorem 4.10 (Eigenspace and Eigenspectrum). For $A \in \mathbb{R}^{n\times n}$, the set of all eigenvectors of A associated with an eigenvalue λ spans a subspace of $\mathbb{R}^n$, which is called the *eigenspace* of A with respect to λ and is denoted by E_λ. The set of all eigenvalues of A is called the *eigenspectrum*, or *spectrum*, of A.

If λ is an eigenvalue of $A \in \mathbb{R}^{n\times n}$, then the corresponding eigenspace E_λ is the solution space of the homogeneous system of linear equations $(A - \lambda I)x = 0$.

Useful properties of eigenvalues and eigenvectors:

- A matrix A and its transpose A^T possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda I$ since

$$\begin{aligned} Ax = \lambda x &\iff Ax - \lambda x = 0 \\ \iff (A - \lambda I)x = 0 &\iff x \in \ker(A - \lambda I) \end{aligned}$$

- Similar matrices possess the same eigenvalues.
- *Symmetric, positive definite matrices* always have positive, real eigenvalues.

Definition 4.11 Let λ_i be an eigenvalue of a square matrix A. Then the geometric multiplicity of λ_i is the number of linearly independent eigenvectors associated with λ_i. It's the dimensionality of the eigenspace spanned by the eigenvectors associated with λ_i.

> A specific eigenvalue's geometric multiplicity must be at least one because every eigenvalue has at least one associated eigenvector.
>
> **Example 4.6**
> The matrix $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has two repeated eigenvalues $\lambda_1 = \lambda_2 = 2$ and an algebraic multiplicity of 2. The eigenvalue has, however, only one distinct unit eigenvector $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and, thus, geometric multiplicity 1.

Theorem 4.12. The eigenvectors $x_1, \ldots, x_n$ of a matrix $A \in \mathbb{R}^{n\times n}$ with n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ are linearly independent.

states that eigenvectors of a matrix with n distinct eigenvalues form a basis of $\mathbb{R}^n$.

Definition 4.13. A square matrix $A \in \mathbb{R}^{n\times n}$ is *defective* if it possesses fewer than n linearly independent eigenvectors.

A *non-defective matrix* $A \in \mathbb{R}^{n\times n}$ require that eigenvectors form a basis of $\mathbb{R}^n$ but not necessarily require n distinct eigenvalues.

The sum of the dimensions of the eigenspaces of. a defective matrix is less than n. Specifically, a defective matrix(has at least one eigenvalue)'s algebraic multiplicity $m > 1$ and geometric multiplicity is less than m.

Theorem 4.14. Given a matrix $A \in \mathbb{R}^{m\times n}$, there's always a symmetric, positive semidefinite matrix $S \in \mathbb{R}^{n\times n}$ by defining

$$S := A^T A$$

If $\text{rk}(A) = n$, then S is symmetric, positive definite.

> *Proof:*

> Understanding why Theorem 4.14 holds is insightful for how we can use symmetrized matrices: Symmetry requires $\boldsymbol{S} = \boldsymbol{S}^\top$, and by inserting (4.36) we obtain $\boldsymbol{S} = \boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}^\top (\boldsymbol{A}^\top)^\top = (\boldsymbol{A}^\top \boldsymbol{A})^\top = \boldsymbol{S}^\top$. Moreover, positive semidefiniteness (Section 3.2.3) requires that $\boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x} \geqslant 0$ and inserting (4.36) we obtain $\boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x} = (\boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{x}) \geqslant 0$, because the dot product computes a sum of squares (which are themselves non-negative).

Theorem 4.15 (Spectral Theorem). If $A \in \mathbb{R}^{n \times n}$ is *symmetric*, there exists an orthonormal basis of the corresponding vector space V consisting of eigenvectors of A, and each eigenvalue is real.

> **Example**:
>
> Consider a matrix
>
> $$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}.$$
>
> The characteristic polynomial of A is
>
> $$p_A(\lambda) = -(\lambda - 1)^2 (\lambda - 7)$$
>
> > computed by $\det(A - \lambda I_3) = 0$
>
> therefore, the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 7$, where λ_1 is a repeated eigenvalue.
>
> From the standard procedure for computing eigenvectors, the eigenspaces are:
>
> $$E_{\lambda_1=1} = \text{span}[\underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}}_{:=x_1}, \underbrace{\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}}_{:=x_2}], E_{\lambda_2=7} = \text{span}[\underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}}_{:=x_3}]$$
>
> > the standard procedure to compute the eigenvectors is $(A - \lambda_i I_N)x = 0$
>
> x_3 is orthogonal to both x_1 and x_2, but x_1 and x_2 is not since $x_1^T x_2 = 1 \neq 0$.
>
> *The spectral theorem* indicates there exists an orthogonal basis, since x_1, x_2 is not, we can construct.
>
> To construct, since x_1, x_2 are eigenvectors share the same eigenvalue λ, for any $\alpha, \beta \in \mathbb{R}$,

$$A(\alpha x_1 + \beta x_2) = Ax_1\alpha + Ax_2\beta = \lambda(\alpha x_1 + \beta x_2)$$

> Use Gram-Schmidt algorithm to find eigenvectors associated with $\lambda_1 = 1$ that are orthogonal to each other (and x_3), so here
>
> $$x_1' = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, x_2' = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

Theorem 4.16. The determinant of a matrix $A \in \mathbb{R}^{n\times n}$ is the product of its eigenvalues, i.e.

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

Theorem 4.17. The trace of a matrix $A \in \mathbb{R}^{n\times n}$ is the sum of its eigenvalues

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

where $\lambda_i \in \mathbb{C}$ are eigenvalues of A.

Cholesky Decomposition

For *symmetric, positive definite matrices*, we can choose from a number of square-root equivalent operations. The Cholesky decomposition/factorisation provides a square-root equivalent operation on it.

Theorem 4.18 (Cholesky Decomposition). A *symmetric, positive definite matrix A can be factorised into a product $A = LL^T$, where L is a lower-triangular matrix with positive diagonal elements:*

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}$$

L is called the Cholesky factor of A, and it's unique.

> 💡 The Cholesky decomposition is an important tool for the numerical computations underlying machine learning.

The covariance matrix of a multivariate Gaussian variable is symmetric, positive definite matrix, so the Cholesky decomposition can help generating samples form a Gaussian distribution.

Given $A = LL^T$, we have $\det(A) = \det(LL^T) = \det(L)\det(L^T) = \det(L)^2$. And since L is a triangular matrix, the determinant is simply the product of its diagonal entries, thus, $\det(A) = \prod_i l_{ii}^2$.

> ### Example

Consider a symmetric, positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{3\times 3}$. We are interested in finding its Cholesky factorization $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} . \quad (4.45)$$

Multiplying out the right-hand side yields

$$\boldsymbol{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix} . \quad (4.46)$$

Comparing the left-hand side of (4.45) and the right-hand side of (4.46) shows that there is a simple pattern in the diagonal elements l_{ii}:

$$l_{11} = \sqrt{a_{11}}\,, \quad l_{22} = \sqrt{a_{22} - l_{21}^2}\,, \quad l_{33} = \sqrt{a_{33} - (l_{31}^2 + l_{32}^2)}\,. \quad (4.47)$$

Similarly for the elements below the diagonal (l_{ij}, where $i > j$), there is also a repeating pattern:

$$l_{21} = \frac{1}{l_{11}} a_{21}\,, \quad l_{31} = \frac{1}{l_{11}} a_{31}\,, \quad l_{32} = \frac{1}{l_{22}} (a_{32} - l_{31}l_{21})\,. \quad (4.48)$$

Thus, we constructed the Cholesky decomposition for any symmetric, positive definite 3×3 matrix. The key realization is that we can backward calculate what the components l_{ij} for the $\boldsymbol{L}$ should be, given the values a_{ij} for $\boldsymbol{A}$ and previously computed values of l_{ij}.

Eigendecomposition and Diagonalisation

💡 **Recap:**

Two matrix A, D are *similar* if there exists an invertible matrix P, such that

$D = P^{-1}AP$.

Diagonalisation

Definition 4.19 (Diagonalisable). A matrix $A \in \mathbb{R}^{n\times n}$ is *diagonalisable* if it's similar to a diagonal matrix, i.e, if there exists an invertible matrix $P \in \mathbb{R}^{n\times n}$ such that $D = P^{-1}AP$.

Let $A \in \mathbb{R}^{n\times n}$, let $\lambda_1, \ldots, \lambda_n$ be a set of scalars, and let $p_1, \ldots, p_n$ be a set of vectors in $\mathbb{R}^n$. Define $P := [p_1, \ldots, p_n]$ and let $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then,

$$AP = PD$$

if and only if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A and $p_1, \ldots, p_n$ are corresponding eigenvectors of A, because

$$AP = A[p_1, \ldots, p_n] = [Ap_1, \ldots, Ap_n]$$
$$PD = [p_1, \ldots, p_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \vdots & \\ 0 & & \lambda_n \end{bmatrix} = [p_1\lambda_1, \ldots, p_n\lambda_n]$$

This requires P is invertible, i.e., full rank.

Eigendecomposition

Theorem 4.20 (Eigendecomposition). A square matrix $A \in \mathbb{R}^{n\times n}$ can be factored into

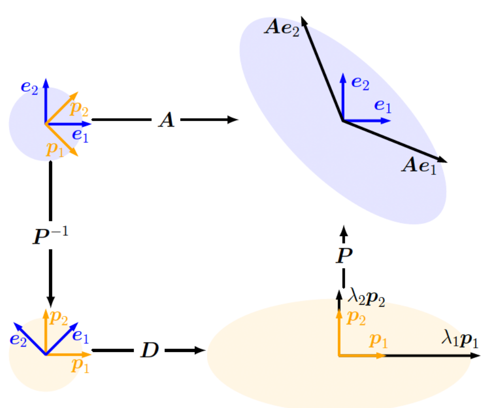
$$A = PDP^{-1}$$

where $P \in \mathbb{R}^{n\times n}$ and D is a diagonal matrix whose diagonal entries are the eigenvalues of A, if and only if the eigenvectors of A form a basis of $\mathbb{R}^n$.

💡 Only non-defective matrices can be diagonalised and that the columns of P are the n eigenvectors of A.

Theorem 4.21. A symmetric matrix $S \in \mathbb{R}^{n\times n}$ can always be *diagonalised*.

An geometric intuitive for eigendecomposition:



> ### Example of eigendecomposition
>
> Compute the eigendecomposition of $A = \frac{1}{2}\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$.
>
> Step 1: Compute eigenvalues and eigenvectors.
>
> The characteristic polynomial of A is

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} \frac{5}{2} - \lambda & -1 \\ -1 & \frac{5}{2} - \lambda \end{bmatrix}\right)$$
$$= (\frac{5}{2} - \lambda)^2 - 1 = \lambda^2 - 5\lambda + \frac{21}{4} = (\lambda - \frac{7}{2})(\lambda - \frac{3}{2}).$$

Therefore, the eigenvalues are $\lambda_1 = \frac{7}{2}$ and $\lambda_2 = \frac{3}{2}$ (the roots of the characteristic polynomial), and the associate (normalised) eigenvectors are

$$Ap_1 = \frac{7}{2}p_1, Ap_2 = \frac{3}{2}p_2$$

and has

$$p_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}, p_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Step2: Check for existence.

The eigenvectors p_1, p_2 can form a basis of $\mathbb{R}^2$, so A can be diagonalised.

Step 3: Construct the matrix P to diagonalise A.

P is the collection of eigenvectors of A

$$P = [p_1, p_2] = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

and then

$$P^{-1}AP = \begin{bmatrix} \frac{7}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} = D$$

- Diagonal matrices D can efficiently be raised to a power.

$$A^k = (PDP^{-1})^k = PD^kP^{-1}$$

- If the eigendecomposition exists, then,

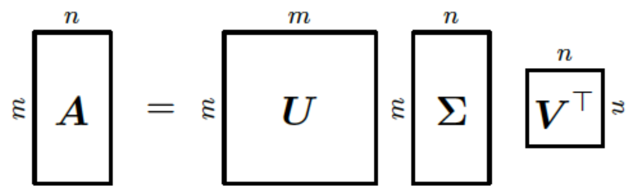
$$\det(A) = \det(PDP^{-1}) = \det(P)\det(D)\det(P^{-1})$$
$$= \det(D) = \prod_i d_{ii}$$

Singular Value Decomposition

The Singular Value Decomposition (SVD) of a matrix is a central matrix decomposition method in linear algebra. SVD can be applied to all matrices, not only to square matrices, and it always exists.

The SVD of a matrix A, which represents a linear mapping $\Phi : V \to W$, quantifies the change between the underlying geometry of these two vector spaces.

Theorem 4.22 (SVD Theorem). Let $A \in \mathbb{R}^{m \times n}$ be a rectangular matrix of rank $r \in [0, \min(m, n)]$. The SVD of A is a decomposition of the form with an *orthogonal matrix* $U \in \mathbb{R}^{m \times m}$ with column vectors $u_i,\ i = 1, \ldots, m$, and an *orthogonal matrix* $V \in \mathbb{R}^{n \times n}$ with column vectors $v_j,\ j = 1, \ldots, n$. Moreover, Σ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij=0}, i \neq j$.



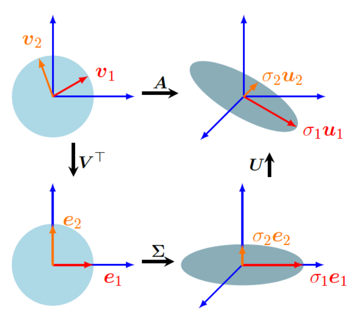
The diagonal entries $\sigma_i,\ i = 1, \ldots, r$, of Σ are called the *singular values*, u_i are called the *left-singular vectors*, and v_j are called the *right-singular vectors*.

The *singular value matrix* Σ is unique.

- $\Sigma \in \mathbb{R}^{m \times m}$ is rectangular and is of the same size as A.
- Σ has a diagonal sub-matrix that contains the singular values and needs additional zero padding.

> 💡 The SVD exists for any matrix $A \in \mathbb{R}^{m \times n}$.

An geometric intuition of SVD transformation.



The SVD expresses a change of basis in both the domain and codomain, which is in contrast with the eigendecomposition that operates within the same vector space.

Construction of the SVD

The SVD of a general matrix shares some similarities with the eigendecomposition of a square matrix.

Computing the SVD of $A \in \mathbb{R}^{m \times n}$ is equivalent to finding two sets of orthonormal bases $U = (u_1, \ldots, u_m)$ and $V = (v_1, \ldots, v_n)$ of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively.

Start with constructing the orthonormal set of right-singular vectors $v_1, \ldots, v_n \in \mathbb{R}^n$, then the orthonormal set of left-singular vector $u_1, \ldots, u_m \in \mathbb{R}^m$. Then link the two and require that the orthogonality of the v_i is preserved under the transformation of A. Since the images Av_i form a set of orthogonal vectors.

Construct the right-singular vecotrs

The spectral theorem (4.15) shows that the eigenvectors of a symmetric matrix form an ONB, which also means it can be diagonalised. Theorem 4.14 shows that we a symmetric, positive semidefinite matrix $A^TA \in \mathbb{R}^{n\times n}$ can always be constructed from any rectangular matrix $A \in \mathbb{R}^{m\times n}$. Thus, diagonalise A^TA

$$A^T = PDP^T = P \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} P^T$$

where P is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geq 0$ are the eigenvalues of A^TA.

Assuming the SVD of A exists

$$A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T$$

where U, V are orthogonal matrices. With $U^TU = I$,

$$A^TA = V\Sigma^T\Sigma V = V \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \vdots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} V^T$$

we identify that $V^T = P^T$ and $\sigma_i^2 = \lambda_i$.

Construct the left-singular vectors U

Similarly, start by computing the SVD of the symmetric matrix $AA^T \in \mathbb{R}^{m\times m}$,

$$\begin{aligned} AA^T &= (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T \\ &= U \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \vdots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} U^T \end{aligned}$$

Link up

Use the fact the images of the v_i under A have to be orthogonal. Require that the inner product between Av_i and Av_j must be 0 for $i \neq j$, For any two orthogonal eigenvectors $v_i, v_j, i \neq j$,

$$(Av_i)^T(Av_j) = v_i^T(A^TA)v_j = v_i^T(\lambda_j v_j) = \lambda_j v_i^T v_j = 0$$

For the case $m \geq r$, it holds that $\{Av_1, \dots, Av_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$.

To complete the SVD construction, the left-singular vectors should. be orthonormal, normalising the images of the right-singular vectors Av_i and obtain

$$u_i := \frac{Av_i}{\|Av_i\|} = \frac{1}{\sqrt{\lambda_i}} Av_i = \frac{1}{\sigma_i} Av_i$$

So the *singular value equation* is

$$Av_i = \sigma_i u_i,\ i = 1, \dots, r$$

or

$$AV = U\Sigma$$

where Σ has the same dimensions as A and a diagonal structure for rows $1, \ldots, r$. Thus $A = U\Sigma V^T$ is the SVD of A.

> **Example:**